DOCUMENT RESUME

ED 167 945 CS 004 564

AUTHOR Tourangeau, Roger; Sternberg, Robert J.
TITLE Aptness in Metaphor. Technical Report No. 13.
INSTITUTION Yale Univ., New Haven, Conn. Dept. of Psychology.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.; Office of Naval Research, Washington, D.C. Psychological Sciences Div.
PUB DATE 30 Oct 78
CONTRACT N0001478C0025
NOTE 63p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Association (Psychological); Associative Learning; *Cognitive Processes; *Comprehension; *Figurative Language; *Metaphors; Reading Comprehension; Reading Research; *Semantics

ABSTRACT

Defining metaphor as "seeing a concept from one class in terms of a concept from another class," a study was devised that analyzed the degree to which two concepts occupy dissimilar positions with respect to their category or domain (within-domain distance), and the degree to which categories themselves are dissimilar (between-domain distance). Preliminary to the study, nine groups of subjects rated the distances among stimuli from eight domains as a basis for two experiments. In the first experiment, one group of subjects rated the aptness of 64 metaphors while another group rated the comprehensibility of the same metaphors. Aptness related positively to between-domain distance and negatively to within-domain distance, but showed no relationship to overall distance. Comprehensibility related strongly to aptness. In the second experiment, subjects ranked a set of alternatives as possible completions of metaphors. In one group, all of the possible completions were from a single domain; in the other group, all were from different domains. In both groups the rank order of the within-domain distances of a set of alternatives correlated with the relative popularity of the alternatives. The overall results of the study consistently supported a domains-interaction view of metaphor. (Tables are appended.) (Author/MAI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED167945

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Aptness in Metaphor

Roger Tourangeau and Robert J. Sternberg

Department of Psychology
Yale University
New Haven, Connecticut 06320

Technical Report No. 13
October, 1978

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

This research was sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-78C0025, Contract Authority Identification Number NR 150-412.

Running Head: Aptness in Metaphor

Send proofs to Roger Tourangeau
Psychology Department
Connecticut College
New London, Conn. 06320

CS004564

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 13	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) "Aptness in Metaphor"		5. TYPE OF REPORT & PERIOD COVERED Periodic Technical Report (1 July 78 - 30 Sept 78) 6. PERFORMING ORG. REPORT NUMBER Research Report No. 13-78
7. AUTHOR(s) Roger Tourangeau and Robert J. Sternberg		8. CONTRACT OR GRANT NUMBER(s) N0001478C0025
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology Yale University New Haven, Connecticut 06520		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N; RR 042-04; RR 042-04-01; NR 150-412
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research (Code 458) Arlington, Virginia 22217		12. REPORT DATE 30 Oct 78 13. NUMBER OF PAGES 60
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) metaphor, domains-interaction theory, within-subspace distance, between-subspace distance, analogy		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The traditional view treats metaphor as a comparison. Because this view has difficulty in explaining the basis of the supposed comparison, the source of the surprise that metaphors often engender, and the role of dissimilarity in metaphor, two other theories have been proposed. One of them treats metaphors as anomalies--intentional mistakes--the theorists differing on the nature of the mistake involved. The other view treats metaphors as interactions, in which one concept is somehow seen in terms of another. We propose		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

a variant of this view: in metaphor, we see a concept from one class in terms of a concept from another class. Our view requires that the similarity (or distance) of the two concepts be analyzed into two components: (a) the degree to which two concepts occupy dissimilar positions with respect to their category or domain (which we call within-domain distance); and (b) the degree to which the categories themselves are dissimilar (which we call between-domain distance).

Each of the views makes predictions about the relations among similarity, aptness, and comprehensibility in metaphor. Aptness or quality, according to the comparison theorists and some anomaly theorists, may be positively affected by overall similarity; or, according to other anomaly theorists, aptness may be adversely affected; or finally, these two effects may combine so that intermediate levels of overall similarity produce the best metaphors. We propose a model based on the distinction between the two components of overall distance; more distance between domains but less distance within domains makes for good metaphors. Two further assumptions shared, in one form or another, by most theorists, are that comprehensibility makes for aptness and that similarity makes for comprehensibility. We undertook several studies to test these positions.

We scaled distances among stimuli from eight domains, using the ratings of nine groups of subjects. The similarity of the dimensional structure of the eight domains supported the notion that dimensions applying in one domain correspond to dimensions in other domains.

In Experiment 1, we obtained ratings of the aptness of 64 metaphors from one group of subjects and ratings of their comprehensibility from another group. Aptness related positively to between-domain distance, negatively to within-domain distance, but not at all to overall distance. Comprehensibility related strongly to aptness.

In Experiment 2, subjects ranked a set of alternatives as possible completions of metaphors. In one group, all the possible completions were from a single domain; in the other group, all were from different domains. In both groups, the rank order of the within-domain distances of a set of alternatives accorded well with the relative popularity of the alternatives. Quantitative predictions, using a choice model patterned after that proposed by Rumelhart and Abrahamson, fared well for the group in which all the choices in a set came from a single domain, but fared very poorly in the group in which all the choices in a set came from different domains.

Overall, the findings consistently supported our domains-interaction view over our versions of the comparison and anomaly views.

Abstract

The traditional view treats metaphor as a comparison. Because this view has difficulty in explaining the basis of the supposed comparison, the source of the surprise that metaphors often engender, and the role of dissimilarity in metaphor, two other theories have been proposed. One of them treats metaphors as anomalies--intentional mistakes--the theorists differing on the nature of the mistake involved. The other view treats metaphors as interactions, in which one concept is somehow seen in terms of another. We propose a variant of this view: in metaphor, we see a concept from one class in terms of a concept from another class. Our view requires that the similarity (or distance) of the two concepts be analyzed into two components: (a) the degree to which two concepts occupy dissimilar positions with respect to their category or domain (which we call within-domain distance); and (b) the degree to which the categories themselves are dissimilar (which we call between-domain distance).

Each of the views makes predictions about the relations among similarity, aptness and comprehensibility in metaphor. Aptness or quality, according to the comparison theorists and some anomaly theorists, may be positively affected by overall similarity; or, according to other anomaly theorists, aptness may be adversely affected; or finally, these two effects may combine so that intermediate levels of overall similarity produce the best metaphors. We propose a model based on the distinction between the two components of overall distance: more distance between domains but less distance within domains makes for good metaphors. Two further assumptions shared, in one form or another, by most theorists, are that comprehensibility makes for

aptness and that similarity makes for comprehensibility.

We undertook several studies to test these positions.

We scaled distances among stimuli from eight domains, using the ratings of nine groups of subjects. The similarity of the dimensional structure of the eight domains supported the notion that dimensions applying in one domain correspond to dimensions in other domains.

In Experiment 1, we obtained ratings of the aptness of 64 metaphors from one group of subjects and ratings of their comprehensibility from another group. Aptness related positively to between-domain distance, negatively to within-domain distance, but not at all to overall distance. Comprehensibility related strongly to aptness.

In Experiment 2, subjects ranked a set of alternatives as possible completions of metaphors. In one group, all the possible completions were from a single domain; in the other group, all were from different domains. In both groups, the rank order of the within-domain distances of a set of alternatives accorded well with the relative popularity of the alternatives. Quantitative predictions, using a choice model patterned after that proposed by Rumelhart and Abrahamson, fared well for the group in which all the choices in a set came from a single domain, but fared very poorly in the group in which all the choices in a set came from different domains.

Overall, the findings consistently supported our domains-interaction view over our versions of the comparison and anomaly views.

Aptness in Metaphor

The traditional view holds that a metaphor is a comparison, a simile stripped of its explicit comparative apparatus. Because the comparative phrasing is left out, the metaphor appears as something other than a comparison, such as a substitution ("anger consumed him") or identification ("jealousy is a green-eyed monster"). Brooke-Rose (1958) lists a number of the grammatical forms that metaphors take; none of them involve comparison overtly, although implicitly, the traditional view holds, all metaphors are in fact making a comparison. Aristotle (Poetics, xxi, 7ff; Rhetoric, III, ii, 7ff.), the source of this comparison theory, also proposed that the terms of a metaphor are members of a common category. Thus, "jealousy" (the subject, or tenor, of the metaphor) and "green-eyed monster" (the object to which the subject is compared, or the vehicle of the metaphor) both belong to some higher-level category, perhaps, in this case, destructive forces. This common category membership reveals the basis of the comparison (the ground of the metaphor). A related view of the ground of metaphor holds that the tenor and vehicle share some characteristics or features; these common features form the ground of the comparison. Almost all theorists--even those who do not believe other aspects of the comparison view--assume that the ground of a metaphor consists of a common category membership or, equivalently, a set of shared features (Basso, 1976; Bickerton, 1969; Chomsky, 1964; Guenther, 1975; Malgady and Johnson, 1976; Tversky, 1977; Van Dijk, 1975; Ortony, in press).

There are two chief rivals to the comparison view. One of these is

the anomaly view, the view that metaphor is intentional or unintentional mistake. The postulated "mistake" derives from the implicitness of the comparison a metaphor makes. Overtly, the metaphor may appear as an identification ("jealousy is a monster"). As an identification, the metaphor is false, even absurd. The obvious dissimilarities between tenor and vehicle render their identification anomalous. If the defining feature of a metaphor for the comparison theorist is the similarity of the tenor and vehicle, then for the anomaly theorist, it is their differences. A number of theorists have proposed anomaly positions on metaphor (e.g., Beardsley, 1962; Bickerton, 1969; Guenther, 1975; Percy, 1954; Van Dijk, 1975; for related views, see Chomsky, 1964; and Katz, 1964); they differ among themselves, however, in their analysis of the postulated anomaly.

The other rival to the comparison view is the interaction view (Black, 1962; Hesse, 1966; Miles, 1967; Richards, 1936; Wheelwright, 1962). If the comparison view stresses the similarity of tenor and vehicle, and the anomaly view stresses the dissimilarity, then the interaction view emphasizes both equally. The interactionists argue that, in a metaphor, the vehicle is a template or model that reshapes our perception of the tenor. This reorganization is necessary, because the characteristics that the tenor and vehicle "share" often aren't really shared at all; the ground of the metaphor may itself be nonliteral. "Men are wolves" (in Black's, 1962, example) because both are predatory, but, the interactionists argue, men and wolves are predators in different ways. There is, at best, only a resemblance between predacity in people and in beasts.

We support a version of the interaction view that we call the domains-

interaction view (see Sternberg, Tourangeau, & Nigro, in press; Tourangeau & Sternberg, Note 1). Domains interactionists assert that in metaphor an object in one category or domain is compared to an object in another category or domain; the domain of the vehicle and the characteristics applying within that domain are the template against which the domain of the tenor is matched. A consequence of this view is that a metaphor often asserts a resemblance between two categories of objects rather than just between two objects. Thus, Miles (1967) observes that:

> What is vital....in metaphor is the sense of relative position within a group or class. If the dove is a cabbage, then the tanager is a carrot.

The interaction that gives the interaction view its name is the conceptual change that results from seeing one thing in terms of another, from seeing one set of characteristics (such as predacity in wolves) as corresponding to another set (aggressiveness in people). This interaction changes our concepts of the tenor and the vehicle. Further, the domains-interaction view assumes that this conceptual change brought about by the metaphor often involves whole domains.

There are several difficulties with the traditional view, particularly with its proposal that a shared category membership forms the ground of metaphor. First, common category membership is often insufficient basis for a comparison--any two things belong to some common superordinate category. Of course, it may well be that most of us can produce some interpretation (i.e., some ground)--however implausible--for any metaphor. But frequently the most obvious category memberships are irrelevant to

the ground. That reporters and badgers, for example, are both animals is plainly irrelevant to the interpretation of "The reporters badgered the politician." In response to this difficulty, some theorists have retained the basic comparison framework, but have assumed that the comparison involves only certain special features or category memberships of the tenor and vehicle. Guenther (1975), for example, argues that only the "prominent" features can form the ground of metaphors. This restriction explains how irrelevant features or categories shared by tenor and vehicle can be excluded from the ground--that reporters are animals (as are badgers) is left out of the ground, since that category membership isn't a prominent one for reporters. The domains-interaction view also suggests a restriction on the sort of feature that can form the ground of a metaphor. The shared features must be among those that give "the sense of relative position within a group." Seeing the ground of the metaphor means seeing two things as having similar positions or roles within their domains. Interaction theorists call attention to a second difficulty, noted earlier, in the comparison theorists' analysis of the ground of metaphor. The common category or feature shared by tenor and vehicle is often only shared metaphorically. The meaning of a feature may differ radically as it applies to the tenor and vehicle: a small headache and a small tornado do not literally share membership in the category of small things.

Another difficulty with the traditional view is that metaphors often involve unfamiliar tenors--tenors for which the category memberships are unknown. Suppose we are describing Ian Paisley to someone who knows only that Paisley is from Northern Ireland; we might say, "Paisley is the George

Wallace of Northern Irish politics." How can the listener find the categories to which Paisley and Wallace both belong, when the listener doesn't know who Paisley is? Yet, it is likely that most listeners could interpret this metaphor, that is, could infer its ground. One solution to this difficulty relies, once again, on the notion that only prominent categories or features (or some other special subset of all the categories or features) can form the ground of the metaphor. Restricting ourselves to the prominent categories, we can infer the ones that the vehicle must share with the unfamiliar tenor--the categories or features that must form the ground of the metaphor. The listener infers that Paisley must be a member of the category "right-wing politician," because that is the only prominent category associated with Wallace to which Paisley might also belong. The domains-interaction view suggests another solution to this problem. According to this view, if we know both the position of the vehicle with respect to its category or domain (that Wallace is on the extreme right of American politics) and the domain of the tenor (that Paisley is a Northern Irish politician), then we can infer the asserted position of even an unfamiliar tenor (that Paisley must also be on the right, relative to other Northern Irish politicians).

Still another difficulty for the comparison view is that metaphors are often surprising. If they rely on extant similarity relations between terms, how can metaphors seem so surprising? Given that we already knew that jealousy is monstrous and destructive, why should Shakespeare's metaphor ever have seemed fresh? A closely related difficulty is that metaphors involve not only similarity, but also the residual disanalogy or dissimilarity

between tenor and vehicle (sometimes called the tension of the metaphor). The comparison view, although it acknowledges the presence of dissimilarity or tension, provides no analysis of its role in metaphor. Surprise and tension, left unexplained in the comparison view, are central to the anomaly view--it is the dissimilarity of tenor and vehicle (i.e., the tension) that produces the anomaly. And it is the anomaly or incongruity of the metaphor's indentification and juxtaposition of diverse elements that produces the surprise in the metaphor. The domains-interaction theory accounts for this novelty by noting that metaphors do not necessarily rely on extant similarities between terms. In order to see objects in different classes as occupying similar positions, we must align the dimensions or features applying within the one category so that they correspond to those applying within the other category. Seeing Wallace and Paisley as occupying analogous positions within their domains requires that the dimension "right wing-left wing," or some analogous dimension, apply to Northern Irish as well as American politics. In this example, the correspondence between the dimensions that give the sense of relative position in the two domains is based on extant knowledge; the correspondence is based on our knowledge that the right-left dimension applies to many political domains. But, in other cases, seeing the correspondence between dimensions requires considerable interpretive activity. In interpreting "men are wolves," for example, we must uncover the correspondence between aggressiveness in men and predacity in wolves. Uncovering this correspondence--aligning these two dimensions--takes interpretive work. The conceptual interaction, the reorganization of our view of the domains, produced by the interpretive

work is, according to the domains-interaction position, central to metaphor. Metaphors, thus, do indeed produce novel and unexpected similarities of relative position within a class--similarities induced by this conceptual interaction--but also involve the perception of the dissimilarity of the classes themselves. Initially, at least, the domains of tenor and vehicle are seen as dissimilar. Having interpreted the metaphor, we may come to see the domains as being similar; but at the outset, the domains seem entirely unrelated.

Besides differing on the nature of the ground, on how we interpret metaphors with unfamiliar tenors, and on the role of dissimilarity and surprise in metaphor, the three general views on the nature of metaphor suggest that different considerations affect aptness in metaphors. For all three positions, a central consideration is the similarity of the tenor and vehicle; the three views differ, however, in their analysis of this similarity.

Hypotheses Regarding Effects of Distance Upon Aptness

Hypotheses Based on Overall Distance

Some theorists taking the anomaly position favor incongruity (e.g., Campbell, 1975; Percy, 1954; see also Wheelwright, 1962, on diaphor) as a source of aptness. For these anomaly theorists, the incongruity of the juxtaposition of diverse elements in metaphor engenders freshness and novelty of vision. Since the chief source of the incongruity (and the resulting novelty) is the overall distance or dissimilarity between tenor and vehicle, then the greater this dissimilarity, the better the metaphor. Other theorists (Chomsky, 1964; Katz, 1964), of a linguistic bent,

emphasize "grammaticality" (whether the sentence is well-formed syntactically) and comprehensibility. For these theorists, the overall dissimilarity of the tenor and vehicle produces grammatical anomaly; in order to overcome this anomaly, special procedures are needed to augment the grammar that applies to "normal" sentences--sentences, that is, that can be interpreted literally. The more serious the deviation, the more elaborate the special procedures. Extrapolating from these views, we might predict that metaphors are worse when the overall distance between tenor and vehicle is greater, because, with increasing dissimilarity, metaphors involve more serious grammatical deviation and so become harder to understand.

For the comparison theorist, the simplest hypothesis is that since a metaphor is a comparison, the closer the comparison (that is, the more similar the tenor and vehicle), the better the metaphor. Although this hypothesis has some empirical support (Malgady and Johnson, 1975), it suffers from a serious theoretical weakness. At some point, tenor and vehicle can be so similar that metaphor degenerates into literal statement of resemblance. "A squirrel has the face of a chipmunk" is hardly a metaphor at all--it certainly is not a good one.

Even Aristotle didn't subscribe to so simple a view as this monotone hypothesis. He advised that "metaphors [vehicles] should be drawn from objects that are related to the object [tenor] in question but not obviously related." Metaphors fail for two reasons: They are too obscure or they are too dull. The overall distance between the tenor and vehicle is positively related to the obscurity of the metaphor but negatively related to the dullness. These two opposed trends may interact, yielding an inverse-U shaped

function relating overall distance and aptness--the best metaphors are then those whose tenor and vehicle are neither too close nor too far from each other.

Hypotheses Based on Two Distances

The domains-interaction position distinguishes two distances: the degree to which two objects occupy dissimilar positions relative to other members of their kind or category (within-domains distance); and the degree of dissimilarity between the domains or categories themselves (between-domains distance). Underlying these two distances are separate sets of dimensions or features. Features or dimensions applying to the domains themselves (between-domains dimensions) determine how close in a higher-order space of domains the two domains are. Features or dimensions applying to the objects within a domain (within-domain dimensions) determine whether two objects in different domains occupy analogous positions within the lower-order spaces of each domain. The relation between the two types of distance is illustrated in Figure 1.

Insert Figure 1 about here

If all the dimensions applying within the one domain correspond to the dimensions applying within the other, and if the values of two objects on corresponding dimensions are identical, then the two objects occupy exactly analogous positions within their domains--within-domain distance is zero. Within-domain distance increases as more within-domain dimensions fail to be aligned with dimensions in the other domain and as the values of the two objects diverge on those dimensions that are seen to correspond. (For a fuller treatment of these concepts, see Tourangeau and Sternberg, Note 1.) We propose, with Miles (1967), that metaphors are better as within-domains distance

decreases: the more the tenor and vehicle occupy analogous positions within their domains, the better the metaphor. Beside within-domain distance, the distance between the domains themselves affects the quality of a metaphor. The novelty produced by the metaphor in part reflects the distance between the domains of the tenor and vehicle. Comparison between objects from close or similar domains (e.g., a land mammal and a bird) are less likely to produce surprise than those between objects from dissimilar domains (a bird and a world leader). We also propose, therefore, that metaphors increase in quality as the distance between the domains of the tenor and vehicle increases. These two hypotheses can be combined:

$$\text{Aptness} = f(d \text{ within}, d \text{ between}) \qquad (1)$$

where f is monotone increasing in distance between domains, and monotone decreasing in distance within domains. Note, then, that in the domains-interaction view, aptness is an "interaction" between these two distances.

As the domains get farther apart, it becomes harder to see the correspondences between the sets of the features applying within the domains. It is easy to see that predacity in birds is like predacity in wolves, but harder to see that aggressiveness in people is also like predacity in wolves. The difficulty of aligning dimensions (seeing that they correspond) within very dissimilar domains suggests that the relation assumed by Equation 1 between aptness and the distance between domains holds only when the domains are close enough so that at least some of the within-domains dimensions can be aligned.

Evidence on the Hypotheses

We have considered four hypotheses concerning similarity and aptness. The positive monotone hypothesis predicts that the aptness of a metaphor

increases as the tenor and vehicle become more similar. This prediction is based either on the assumption that comprehension is easier when tenor and vehicle are more similar (as some anomaly theorists might argue) or on the assumption that metaphors are comparisons (as comparison theorists argue). Other anomaly theorists, stressing novelty rather than clarity, offer the opposite prediction, the negative monotone hypothesis: Metaphors are more apt as the objects they link become less similar. These two opposed trends may combine to yield a third relation, a nonmonotone inverse-U relation: The best metaphors involve intermediate similarity. Finally, the domains-interaction view suggests that aptness increases with distance between domains but decreases with distance within domains.

There is little evidence to decide among the general positions on metaphor and the hypotheses regarding similarity and aptness derived from them (see Henle, 1962, for an early review; Billow, 1977; Ortony, Reynolds, and Arter, 1978; and Tourangeau and Sternberg, Note 1, present more recent reviews). Two studies (Malgady & Johnson, 1976; Kraut, Note 2) have examined the relationship between overall similarity and quality, both with somewhat inconsistent results. Neither study distinguished distance between domains from distance within domains--a crucial distinction for the domains-interaction theory we espouse.

We, therefore, undertook several studies to test the hypotheses concerning aptness in metaphor. Eight groups of pretest subjects rated objects from eight domains; a ninth group rated the domains themselves. We derived measures of overall, within-domains, and between-domains distance from these ratings. In Experiment 1, one group of subjects (1A) rated

metaphors on scales of aptness; another group (1B) rated these metaphors on scales of comprehensibility. In Experiment 2, subjects ordered possible ways to complete metaphors that lacked a vehicle. For one group of subjects (2A), all the possible vehicles were members of a single domain; for another group (2B), the vehicles were members of different domains.

Scaling Prestudies

The purpose of the scaling prestudies was to test our assumption that dimensions applying within a domain have analogues in other domains. Further, we needed to scale the distance relations among a set of conceptual objects in order to test our hypotheses concerning distance and aptness in metaphor.

Method

Subjects. One hundred sixty subjects participated in a study of "Similarity Relations Among Concepts." Some subjects were recruited from the introductory psychology course at Yale University and received credit for participating; the rest were recruited via posters and paid $2.50. The subjects participated individually or in groups of from two to six members.

Materials. Each subject received a test booklet; for eight groups (n=16) of subjects, the booklets contained twenty items from a single domain; for the ninth group (n=32), the booklets contained only eight items (the names of the eight domains). The domains were aquatic animals, birds, land mammals, ships, aircraft, land vehicles, U.S. historical figures, and modern world leaders. These domains were selected because we believed that subjects would recognize many objects within each domain, and that at least three common dimensions would apply within each: power, aggression, and prestige.

Following each item in the booklet were 21 seven-point scales with labeled endpoints. A scale looked like this:

Powerful 1 2 3 4 5 6 7 Weak.

Subjects indicated their judgments by circling the most appropriate number. We included seven scales seemingly related to power, seven seemingly related to aggression, and seven seemingly related to prestige. For each item, the 21 scales were in different random orders; likewise, for each subject, the items were in a different random order.

In summary, eight groups of subjects received booklets, containing twenty items (in a random order), each followed by 21 rating scales (in different random orders for each item). The ninth group had booklets containing only eight items (the eight domains), followed by the 21 scales. Subjects in this group also rated the similarity of all 28 possible pairs of domains on a nine-point scale with values ranging from "Like" to "Unlike." The similarity judgments preceded the rating of the domains on the 21 rating scales.

Procedure. We told subjects to rate all the items on every scale and to work rapidly, basing their ratings on their initial impression. In order to encourage their use of the full range of scale values, we first asked subjects to "thumb through the booklets and get an idea of the set of things you'll be rating," since the judgments were to be made relative to the set of objects rated. Subjects were told that if they didn't recognize an item, they should ask the experimenter to describe it for them.

Results

For all nine sets of ratings, we obtained the mean rating (averaged

across subjects) and variance of each item on all the scales.

For the eight sets of ratings of the items within domains, we dropped the three most variable scales, thereby excluding one scale from each of our three a priori clusters of scales. In each of the domains, we found the correlations between all possible pairs of the 18 scales. Each correlation in this 18 x 18 matrix was based on 20 observations, with each observation the average rating of one of the 20 items on a scale. Thus, even though each correlation was based on a small number of observations, we nevertheless hoped the correlations would be reliable because each observation, being an average based on 16 subjects, would be quite stable. The reliability of the observations (split across arbitrary halves of subjects and corrected by the Spearman-Brown formula) was .86.

We analyzed these correlations by principal-axis factor analyses, and rotated all solutions to a varimax criterion. In all eight domains, although our three a priori clusters were still apparent, only two factors emerged. One of the factors clearly reflected prestige; in all eight domains the six scales relating to prestige loaded near 1.0 on this factor and near 0.0 on the other factor. The other factor reflected power or aggression. Table 1 gives the factor loadings for the domain of birds. These loadings are typical of the other domains.

Insert Table 1 about here

Table 2 presents correlations between loadings of corresponding factors in the eight domains. Correlations between loadings of the noncorresponding factors were generally trivial. It can be seen that the correlations are

generally very high, indicating that the factor spaces are quite similar. The median r between corresponding factors is .93. The factor spaces are so similar that we did not apply any technique of rotation to maximum congruence ("Procrustean rotation") to improve the fit.

Insert Table 2 about here

We formed approximate factor scores for each item by adding average ratings of that item on those scales that were "salient" on the factor; we considered a scale salient on a factor when the scale loaded highly on that factor and trivially on the other factor (see Gorsuch, 1974). This criterion meant that we formed the prestige factor score, for items in all eight domains, by adding the item's mean ratings on the six scales in the prestige cluster. We used the same procedure in forming the power-aggression factor scores. These approximate factor scores served as the coordinates (on within-domain dimensions) of items in the main studies.

We took the mean ratings of the domains themselves on the 21 scales and, treating the domains as variables, conducted a principal-axis factor analysis. Three factors emerged. On one, the three animal domains (birds, aquatic animals, land mammals) load highly; on another, the three vehicle domains (aircraft, ships, land vehicles) load highly; on the third, the two domains of people (U.S. figures and world leaders) load highly. The loadings are shown in Table 3. The factor loadings of the domains served as the coordinates of the between-domain dimensions in the main studies.

Insert Table 3 about here

The mean ratings of the similarity of pairs of domains were scaled using nonmetric multidimensional scaling (Kruskal, 1964a, 1964b; Shepard, 1962a, 1962b). Three-dimensional solutions, although in line with the results of the factor analysis, appeared "degenerate," the eight domains collapsing into four clusters: (a) U.S. figures and world leaders; (b) aircraft and birds; (c) ships and land vehicles; and (d) aquatic creatures and land mammals. These solutions were not further used.

Experiment 1

The similarity of the dimensions found in the eight domains lent some support to our assumption that corresponding (or common) dimensions apply within different categories or domains. The dimension that distinguishes Lincoln, Grant, and other war presidents from Thomas Jefferson and Grover Cleveland, for example, corresponds to the dimensions that distinguish battleships from ocean liners and hawks from nightingales.

The purpose of Experiment 1 was to compare the several hypotheses concerning similarity and aptness. Having analyzed the prestudy data, we were ready to derive measures of psychological distance. In Experiment 1A, we collected ratings of the aptness of 64 metaphors. Further, since several of the hypotheses depend on considerations of comprehensibility, we collected ratings of the comprehensibility of the metaphors in a second study, Experiment 1B. Finally, we sought to test the implications of the several positions for individual differences: Do naive and sophisticated readers differ in their rating patterns? We administered tests that were relevant, we hoped, to distinguishing naive readers from sophisticated ones.

The basic approach to distance in all of the studies was dimensional and Euclidean. The dimensional approach treats a concept as a point in a semantic or conceptual space; the point's coordinates give its location (i.e., the concept's value) on the salient dimensions of meaning. The degree to which two concepts differ--their dissimilarity--is measured by the Euclidean distance between them.

Our hypotheses involved three different sorts of dissimilarity or distance. One is overall distance. The overall distance between two concepts is a function of their differences on all relevant dimensions (even those that, according to the domains-interaction viewpoint, apply to the domain of the concept, rather than to the concept itself). Our analyses of the prestudy data uncovered a total of five dimensions (viz., the two within-domain dimensions--power/aggression and prestige--and the three between-domain dimensions--animals, humans, and vehicles). We assigned each concept five coordinates.[1] Overall distance, then, is

$$d(A,B) = [\sum_{i=1}^{n}(a_i - b_i)^2]^{\frac{1}{2}} \qquad (2)$$

The overall distance between two concepts (A and B) is a function of the difference between their coordinates (a_i and b_i) on the n (here five) dimensions.

The other two sorts of distance were also derived using the Euclidean function given in Equation (2). For within-domain distance, however, the summation only ranged over the within-domain dimensions; for between-domain distance, the summation ranged over the between-domain dimensions. Here, the concepts, A and B, with coordinates a_i and b_i, were two domains. The definition of within-domain distance is, thus, the same when the two concepts belong to the same domain as when they belong to different domains--

it is only their relative positions within a domain that affect the within-domain distance between two concepts.

Having derived these measures of overall distance, distance between domains, and distance within domains, we tested several hypotheses against the aptness ratings obtained in Experiment 1A--that aptness is negatively related to overall distance; that it is a nonmonotonic, inverse-U function of overall distance; or finally, that aptness is positively related to the distance between the domains of the tenor and vehicle, but is negatively related to the distance within th -espective domains. Further, we tested several hypotheses concerning comprehensibility. These hypotheses are assumed in the predictions relating aptness and distance -- that comprehensibility varies inversely with overall distance, that it varies inversely with the distance between domains and with the distance within domains, and that it is positively related to aptness.

Method

Subjects. The 37 subjects in Experiment 1A were recruited from the introductory psychology course and received credit toward fulfillment of a course requirement. The 20 subjects in Experiment 1B were students and others in the New Haven area. They were recruited by posters and paid $2.50. The subjects participated individually or in groups of from two to six members.

Materials. Subjects in both groups filled out booklets containing 64 metaphors of the form, "The owl is the horse among birds." Although all the subjects received the same 64 metaphors, each one got them in a different random order. Four scales for rating aptness (in Experiment 1A)

or two scales for rating comprehensibility (in Experiment 1B) followed each metaphor.

The metaphors were selected by pairing tenors from each domain with vehicles from each domain; all 64 possible permutations of domains were represented.[2]

Rating scales. Subjects in both groups rated the metaphors on nine-point scales with labeled endpoints by circling the number they thought most appropriate.

Subjects in Experiment 1A rated the metaphors on four scales, labeled "Good-Bad,""Apt-Not Apt," "Interesting-Dull," and "Like-Dislike," with higher numbers indicating better metaphors. Subjects in Experiment 1B rated the metaphors on two scales, labeled "Hard-Easy" (with higher numbers indicating metaphors that are easier to understand) and "Slow-Fast" (with higher numbers indicating metaphors that are understood more quickly).

Individual difference measures. We hoped to tap individual differences in literary experience and included a measure for this purpose. Further, we suspected that the key advantage of more experienced readers lay in their ability to see the correspondences between dimensions in diverse domains. We included a set of analogy problems in an attempt to detect differences in this ability. As it turned out, neither measure proved reliable; nor did either relate to any of the other variables. For these reasons, we will not consider these measures any further.

Procedures. Subjects were told that the experiment involved aptness (Experiment 1A) or comprehensibility (Experiment 1B) of metaphor. The subjects in both groups received oral and written instructions explaining the use and

meaning of the rating scales. We told the subjects to use the full range of scale values in rating the metaphors, emphasizing, for the subjects in Experiment 1A, that they should use the full range of values, even if none of the metaphors seemed to them, in an absolute sense, to be good ones. We encouraged subjects to ask about any items they didn't recognize in the booklet of metaphors (and, indeed, some of the terms were unfamiliar to the subjects). The experiments lasted about an hour.

Results

We found for each of the 64 metaphors its mean (over subjects) on the rating scales of aptness and comprehensibility. These overall means (or means based on subgroups of subjects) were treated as the basic data in the analyses.

Aptness. The four scales of aptness proved highly interrelated. We summed these scales into a single overall scale, labeled "quality" in Table 4. Correlations between pairs of scales are shown in Table 4. Except as noted, the pattern of results for this composite held also for

Insert Table 4 about here

the individual scales. Both between-domain and within-domain distance related significantly to quality: the correlation between quality of a metaphor and the distance between the domains of its tenor and vehicle was .27; the correlation between quality and within-domain distance was -.39. In line with the domains-interaction hypothesis, metaphors got worse as the within-domain distance between tenor and vehicle increased, but got better as the distance between their domains increased. Further, overall distance

did not significantly relate to quality. The effect of the dimensions on which distance related positively to quality apparently cancelled that of the dimensions on which distance relates negatively to quality. These results are shown in Table 5. As Table 6 shows, distance on both of the within-domain dimensions related negatively to quality, while distance on

Insert Table 5 about here

two (the human and animal dimensions) of the three between-domain dimensions related positively.

Insert Table 6 about here

Several of the hypotheses involved inverse-U shaped functions relating distance and aptness. We looked for such trends by examining scatter-plots and by computing correlations between squared overall and between-domain distances (with these distances first standardized to give them a mean of zero). Neither method revealed any evidence for an inverse-U shape trend.

For all the scales except "interest," the distance of the tenor and vehicle within their domains was more strongly related to aptness than was the distance between domains. That the relation of aptness to between-domain distance was weak is more apparent when the eight metaphors whose tenor and vehicle were from the same domain are removed from consideration; when these eight extreme cases (in which between-domain distance is zero) were excluded, the correlations between the measures of quality (except for the "interest"scale) and distance between domains were no longer significant, although they remained negative, as shown in Table 7. According to the

domains-interaction view, these extreme-case metaphors are anomalous; according to the ratings, they make poor metaphors.

Insert Table 7 about here

Comprehensibility. Several of the hypotheses are based on the idea that increasing distance makes comprehension harder. Since the rated speed and ease of comprehension were highly related (r = .82), we summed them to give a single comprehensibility measure. We could find no evidence in support of the hypothesis that distance and comprehensibility are related; overall, between-domains, and within-domains distance did not relate significantly to the scale of comprehensibility. None of the correlations exceeded .10.

Other factors, such as the familiarity of the terms involved, may have affected comprehensibility so much that the effects of distance were hard to tease out. Presumably, to understand a metaphor we must locate the tenor and vehicle (or, at least, the vehicle) within their domains; locating the concepts requires, further, that we know the domains to which they belong, the dimensions that are important within those domains, and the coordinate values of the concepts on those dimensions. Since the domain of the vehicle serves as a kind of template for viewing the domain of the tenor, it is particularly important that we know the dimensions salient for the vehicle.

One property of the vehicle that may make it easier to infer the dimensions relevant within its domain is the extremity of the vehicle: the more extreme the concept, the more salient the dimensions are on which it is extreme. We formed a measure of the extremity of the tenor and vehicle

by summing the absolute values of their (standardized) within-domain coordinates. The extremity of the vehicle, but not that of the tenor, correlated significantly with the measure of comprehensibility (r = .26 for the extremity of the vehicle, p <.05; r = .06 for that of the tenor).

If the tenor is to be viewed in terms of the vehicle, it may help for the tenor to be a source of some disagreement or uncertainty, but for the vehicle to be unambiguous; if we want to bring the tenor and its domain into correspondence with that of the vehicle, it may make the alignment easier if we are uncertain about the tenor or if we already know that several views of the tenor are possible. We formed a measure of the variability of the tenor and vehicle, by summing their standard deviations on the scales composing the within-domain coordinates. We hoped the measure would serve as an index of the flexibility with which the concept could be viewed. As predicted, the variability of the tenor, but not that of the vehicle, correlated significantly with the measure of comprehensibility: for the tenor, the correlation was .31 (p <,05); for the vehicle, the correlation was .12.

Comprehensibility and aptness. Whatever its determinants (the familiarity of tenor and vehicle, their extremity, the flexibility with which we can view them), comprehensibility is related to aptness or quality. Both rating scales of comprehensibility and their sum related significantly to the measures of quality, as shown in Table 8. These effects were

Insert Table 8 about here

independent of those involving within and between-domain distance; comprehensibility did not interact with either form of distance in its relation

to the measures of quality, nor did the strength of the relations between the comprehensibility scales and the scales of aptness diminish when the effects of distance were partialled out.

Taken together, the comprehensibility of a metaphor, the distance within domains between its tenor and vehicle, and the distance between the domains themselves predicted most of the reliable variation in the metaphor's quality. We formed a single index of comprehensibility by summing the ease and slowness scales. When this measure of comprehensibility was combined with distance within and between domains in a single regression equation, the multiple correlation with quality was .76 ($F(3,60) = 27.3$, $p < .01$). As Table 9 shows, all three variables make a significant contribution. This degree of prediction is more impressive given the limited reliability

Insert Table 9 about here

of the predictors.

Experiment 2

In Experiment 1, we tested hypotheses concerning relations between similarity, in its several forms, and aptness, with rating scales serving as measures of aptness. In Experiment 2, we sought to extend these results to a different mode of response, the selection of vehicles for incomplete metaphors. This response mode had the advantage that it allowed us to explore the relation between the domains-interaction position on metaphor and Rumelhart and Abrahamson's (1973) theory of analogical reasoning.

Rumelhart and Abrahamson claimed that the answer we choose to complete a standard analogy problem (of the form A:B::C:-----) is a function of an

ideal concept--the concept that bears exactly the same relation to the C term of the analogy as the B term bears to the A term. Using a spatial representation, Rumelhart and Abrahamson represented A and B as points in a conceptual space; the relation between A and B was represented by a vector connecting them. If the vector is moved so that C becomes its origin (instead of A), then the point at which it terminates represents the ideal concept.

Our model of the process of selecting vehicles to complete metaphors is patterned after Rumelhart and Abrahamson's model for analogies. Given our assumption that a concept's characteristics are defined in relation to its domain, we can view the tenor or vehicle not only as a point in a domain, but also as a vector from the origin of the domain. This vector reflects the relation of the concept to its domain. Having aligned the domain of the vehicle with that of the tenor, we find an ideal vehicle (the concept whose position within its domain is exactly analogous to that of the tenor) by moving the vector representing the tenor to have as its origin the origin of the domain of the vehicle.

In their mathematical model of the response-selection process, Rumelhart and Abrahamson used the Luce (1959) choice rule. The rule assumes that the probability of selecting an answer is a function of the answer's value, relative to the values of all the alternatives in the choice set:

$$p(A_j) = v(A_j)/\Sigma_i v(A_i). \qquad (3)$$

The probability of choosing one alternative (A_j) from a set is the value of that alternative over the sum of the values of all the alternatives.

Rumelhart and Abrahamson assumed that the "value" of an analogy answer is a decreasing, exponential function of its distance from the ideal solution:

$$V(A_j) = \exp(-\alpha \cdot \text{distance}_A), \qquad (4)$$

where α is the "slope" of the exponential. Intuitively, Equation (4) states that an alternative becomes less attractive, the further it is from the ideal concept, and that differences among bad alternatives matter less than differences among better ones.

Rumelhart and Abrahamson's subjects ranked all the alternatives, rather than selecting just their most favored one. Rumelhart and Abrahamson extended the Luce choice rule to ranking by assuming that subjects apply it repeatedly: subjects make their first choice in accordance with the choice rule, then decide among the remaining alternatives, again choosing the best remaining alternative in accordance with the choice rule; and so on, until all the options are ranked.

How might we predict subjects' choices among vehicles to complete a metaphor? The direct analogue to Rumelhart and Abrahamson's Equation (4) is one based on the within-domain distance of the vehicle from the tenor; the "ideal" vehicle in a domain is that concept whose within-domain distance from the tenor is zero. We can, therefore, replace distance from the ideal in Equation (4) with distance within domains. This model, however, ignores the distance between the domains of the tenor and vehicle--and this distance also contributes (albeit weakly) to the quality of the metaphor. Equation (5) corrects this omission by assuming that the value of an alternative is a function of its aptness:

$$V(A) = f(\text{aptness}), \qquad (5)$$

where f is nonnegative and monotone increasing in aptness. If we assume that good alternatives are more finely discriminated than poor ones, we can predict, with Rumelhart and Abrahamson, that value to be an exponential function.

$$V(A) = \exp(\alpha \cdot \text{aptness}_A). \qquad (6)$$

Method

Subjects. Twenty subjects participated in each group of Experiment 2. Subjects in Experiment 2A were recruited from the introductory psychology course at Yale and received credit toward fulfillment of a course requirement. Subjects in Experiment 2B (most of whom were students) were recruited by poster and paid $2.50. The subjects participated individually or in groups no larger than six.

Materials. The subjects in both groups of Experiment 2 received booklets containing 32 metaphors with missing vehicles: "The bomber is the----- among things that fly." Four possible vehicles followed each metaphor. Although all the subjects received the same 32 incomplete metaphors, each got them in a different random order. The tenors of these metaphors were chosen by selecting an item at random from each quadrant of each of the eight domains. For the subjects in Experiment 2A, the four possible vehicles following each metaphor were from a single (randomly selected) domain, with each vehicle drawn (at random) from different quadrant of that domain. For the subjects in Experiment 2B, the four possible vehicles were from four different domains (chosen at random), with each as close as possible to the ideal vehicle in its domain: within its own domain, that is, each of these vehicles had the smallest within-domain distance from the tenor.

Procedure. Subjects in both groups received booklets containing the 32 metaphors. The subjects were told to rank the four choices according to how well each completed the metaphor, and they were encouraged to ask the experimenter about items they did not recognize. Subjects were given a half hour to complete the booklet. (They also filled out the two ability measures used in Experiment 1.)

Results

For each set of four possible vehicles, we found the proportion of subjects who ranked it best, second best, and so on. The analyses below treat these proportions as the unit of observation.

Qualitative results. The most basic predictions of the several positions are ordinal: more subjects will select as best a vehicle that is closer within domains (or closer overall) than one that is farther; more subjects will choose an alternative vehicle from a domain that is farther from that of the tenor than from one that is closer to the domain of the tenor. These ordinal predictions do not depend on the form of the decision rule that subjects use; nor do they make any assumptions about the process by which subjects make subsequent choices after their first one.

We ranked each alternative vehicle in a set of four according to its within-domain distance and its popularity as the vehicle ranked best. The correlation of these ranks over the 32 metaphors in Experiment 2A was -.46 (based on the 96 independent ranks); in Experiment 2B, this correlation was -.48 (again, based on the ranks of three vehicles for each of the 32 metaphors). In Experiment 2B, the alternatives in each set varied in

their between-domain, as well as their within-domain, distance from the tenor. The correlation of the rank between-domain distance and the rank popularity of an alternative was only .06. In Experiment 2B, we could compare the rank overall distance with rank within-domain distance. (In Experiment 2A, these ranks correlated perfectly because all the vehicles in a set were from a single domain; within a set of alternatives, the only sources of variation in overall distance were on the two dimensions that also determined within-domain distance.) The correlation of a vehicle's rank overall distance within a set of alternatives and its rank popularity among these alternatives was -.04. Thus, in both experiments, the rank within-domain distance of a vehicle significantly predicted its rank popularity within a set of alternatives; and in Experiment 2B, where comparison was possible, it predicted better than rank overall distance (Hotelling's $t = 5.24$; $p<.002$, two-tailed; see Guilford, 1956, page 190).

Quantitative models. For the first-choice data of both experiments, we estimated with nonlinear regression techniques (using the BMD program for nonlinear regression) the best values for the slope parameters (α) of several, related exponential functions:

$$P(A_i) = \exp(\alpha \cdot x_i)/\Sigma\exp(-\alpha \cdot x_i). \qquad (7)$$

In one model, we replaced x_i with the predicted aptness of a vehicle (A_i); in another, we replaced x_i with the vehicle's within-domain distance from the tenor; in another, we used the overall distance; in Experiment 2B, we were able to use the between-domain distance. The predicted aptness of a vehicle is a linear combination of its within-domain and between-domain distance; we used the weights and intercept estimated in Experiment 1.

For the models based on aptness and distance between domains, in which the predictor was positively related to an alternative's attractiveness we expected α to become negative.

Since subjects ranked all four of the possible vehicles from best to worst, there were four proportions associated with each one (viz., the proportion ranking it best, the proportion ranking it next best, etc.) and thus a total of nine degrees of freedom for each metaphor. Using the several models based on equation 7 (with estimates of α for each model estimated from the first choice data), and applying these models successively to all four ranks, we made predictions concerning all nine independent proportions for each of the 32 metaphors.

In Experiment 2, the simple correlation between the predicted and obtained proportions was highest for the model using aptness as the predictor (x_i) in Equation (7), $\underline{r}(286) = .44$, $\underline{p}<.001$, two-tailed. But since the between-domain distance of the alternatives in a set did not vary in this experiment, predicted aptness was confounded with an alternative's within-domain distance from the tenor and (to a lesser extent) its overall distance. It came as no surprise, therefore, that the model using the alternative's within-domain distance (instead of its predicted aptness) as the predictor in Equation (7) did as well ($\underline{r} = .43$), or that the alternative's overall distance was only slightly less successful as the predictor ($\underline{r} = .37$). And a linear model--

$$P(A_j) = \text{Aptness}_{A_j} / \Sigma_i \text{Aptness}_{A_i} \qquad (8)$$

that uses parameters estimated from Experiment 1 (and no new ones estimated from the data of this experiment) also did nearly as well as these exponential

models (r = .36).

For each metaphor, we arranged the four vehicles in order of their popularity as predicted by each model. We then averaged the predicted and obtained proportions across all the metaphors. Figure 2 shows these overall predicted and obtained proportions for the exponential model with aptness as the predictor. The fit appears substantial (r = .98); the departures from prediction are, however, highly significant, $\chi^2(9) = 23.5$, $p < .01$.

In Experiment 2B, none of the correlations between the proportions predicted by the quantitative models and those observed was significant. The "best" model was again the exponential model with aptness as the predictor, but the correlation between predicted and observed proportions was a mere .022. Given the significant qualitative fit of the within-domain distance with the first-choice ranks of this experiment and the relative success of the quantitative models in Experiment 2A, this poor quantitative prediction is puzzling. In selecting the stimuli for this experiment, we tried to minimize variations among the alternatives in a set in their within-domain distance from the tenor. Since within-domain distance (of all the types of distance) was the most consistent predictor of aptness, perhaps restricting its range guaranteed poor quantitative prediction. Or, since the subjects in Experiment 2B were drawn from a different population from those in 2A, it is possible subjects used a decision procedure that is not well described by the Luce choice rule, or one that is insensitive to the spacing among the alternatives.

Discussion

Findings Bearing on the Main Hypotheses

In the scaling prestudies, we scaled stimuli from several domains, using factor analytic techniques. The domains-interaction position assumes that analogous dimensions apply within different domains. Such correspondences are clearly observed in the domains studied here: two factors emerged within each domain and the correlations across domains of the corresponding factor loadings were very high (see Table 2).

The main hypotheses, however, concerned the relation between similarity of the tenor and vehicle (defined in several ways) and the aptness of the metaphor that links tenor and vehicle. The comparison view suggests most simply that aptness varies directly with overall similarity; the views of some anomaly theorists also suggest this relationship. In either case, the prediction seems to rest on two assumptions: First, more overall similarity between tenor and vehicle increases the clarity or comprehensibility of a metaphor; and, second, comprehensibility positively relates to aptness. Some anomaly theorists emphasize novelty, instead of clarity, as a determinant of aptness. These theorists predict a negative relation between aptness and overall similarity. These two trends may combine to produce an inverse-U shaped relation between overall similarity and aptness. Although he did not put it quite this way, Aristotle seems to have supported this nonmonotonic hypothesis.

The evidence of Experiments 1 and 2 provided scant support for any of the proposed relations between overall similarity and either aptness or comprehensibility: None of the relevant correlations were significant.

In Experiment 2, the overall similarity of a vehicle to its tenor predicted the popularity of the vehicle only when overall similarity was confounded with within-domain similarity (in Experiment 2A). Overall similarity bore no significant relation to measures of comprehensibility either. Only the assumption that comprehensibility relates positively to aptness appeared to be justified; measures of the two were significantly related in Experiment 1.

The domains-interaction view distinguishes the within-domain distance between two objects from their between-domain distance: The within-domain distance depends upon the extent to which two objects occupy analagous positions within their domains; the between-domain distance depends on the similarity of the domains themselves. According to the domains-interaction view, within-domain distance relates negatively to aptness, but between-domain distance relates positively. Both of these predictions received some support in Experiment 1: Both within-domain and between-domain distance related significantly to aptness in the expected directions. In Experiment 2, however, only within-domain distance seemed to matter. In both groups of Experiment 2, qualitative predictions based on a vehicle's within-domain distance from the tenor related to its rank popularity as the best completion of a metaphor; and in the first group (but not in the second), quantitative models (patterned after the one proposed by Rumelhart and Abrahamson) based on within-domain distance also significantly predicted the popularity of an option. In contrast, the between-domain distance afforded neither qualitative nor quantative prediction in the one study, 2B, where it varied.

Thus, although between-domain distance is a significant predictor in Experiment 1, it is a poor predictor in Experiment 2. This difference between the results of the two studies may reflect the difference in the dependent measures. Experiment 1 forced subjects to consider not only the aptness of a metaphor but also the interest--and interest, unlike the other measures of Experiment 1, mainly reflected distance between domains. Perhaps because the subjects explicitly considered interest, their ratings on the other scales also reflected the distance between the domains of the tenor and vehicle. Nothing in Experiment 2 forced the subjects to consider interest, surprise, novelty, or anything else determined mainly by distance between domains; subjects in Experiment 2 were simply to rank the choices according to how <u>well</u> they completed the metaphor. Another difference between the studies was that ranking the choices is harder than making direct ratings. In order to reduce this difficulty, subjects in Experiment 2 may have used a simpler decision rule than subjects in Experiment 1, a rule that reflected only one factor, the distance within domains.

<u>Additional Findings</u>

The interaction position on metaphor has as its central assumption the view that, in interpreting a metaphor, we see the tenor in terms of the vehicle. A consequence of this assumption is that tenor and vehicle have asymmetrical roles in the interpretive process. In Experiment 2B, we found some (albeit weak) evidence for this asymmetry. The interpretive process, we assume, requires that we know the relevant characteristics of the vehicle, since they are to form the template for seeing the tenor, and that we are

able to view the tenor with enough flexibility that we can see it in this new way. If we assume that extreme values in the vehicle help us find its relevant characteristics by making them more salient, then extremity should facilitate comprehension. Our finding that the extremity of the vehicle, but not that of the tenor, correlated with comprehensibility, lent some support to this view. If we also assume that disagreement about a concept's position indicates there is some flexibility in the way it can be seen, then measures of disagreement about the tenor's location (such as the variability of its coordinates) should relate to comprehensibility. We also found some evidence that the variability of the tenor, but not that of the vehicle, related to comprehensibility.

The domains-interaction view asserts that a major component of the interpretive process requires "aligning," or seeing the correspondence between analogous dimensions applying in different domains. We assumed that the similarity of the domains affects the ease with which we perform this alignment and thereby see the correspondences between analogous dimensions. Based on this reasoning, we predicted the distance between domains would relate negatively to comprehensibility. This prediction received no support from the relevant data (in Experiment 1B).

Limitations on the Studies

The purpose of the experiments was not to account for all of the reliable variation in the quality or comprehensibility of metaphors. Instead, the purpose was to test the implications of three general views on metaphor by examining the relations among distance (defined in several ways), comprehensibility, and aptness.

Indeed, it would have been quite a surprise if we had been able to predict the aptness or comprehensibility of the metaphors perfectly. In the first place, the use of a spatial representation, whatever its other virtues and limitations, requires, for practical reasons, that only a few salient dimensions be included; in these experiments, the crucial within-domain distance measure is based on only two dimensions. But the concepts involved have many more properties. Almost every metaphor we used involved some extra point of comparison, unrelated to either of our within-domain dimensions. So it seems unlikely we measured similarity exhaustively.

In the second place, there are purely technical limitations on the possible degree of relation among the variables. Some of the measures, including within-domain distance, were not particularly reliable. Further, more than a year passed between the collection of the scaling and aptness data; in the meantime, perceptions of some of the items, especially the political figures, doubtless changed. The metaphors, particularly those in Experiment 1, were chosen to be unbiased, not to be good. As a result, the range of their quality was somewhat limited. In Experiment 2, only similarity (and not comprehensibility) was included as a predictor, although all the main positions acknowledge the effect of comprehensibility. Given these sources of inaccuracy and incompleteness, what is perhaps surprising is the degree of prediction attained. In Experiment 1, for example, comprehensibility, within-domain, and between-domain distance, seem almost to have reached the limit (given the reliabilities of the measures) in the prediction of aptness.

There are other limitations on the methods of these studies, besides

the technical limitations and those inherent in the use of a spatial representation of meaning. Prediction required a number of assumptions, unrelated to the main hypotheses of any of the positions. Although the techniques we used that depend on these assumptions were to a degree robust (giving accurate results even when there are small violations of the assumptions), they were used for convenience' sake, and, at best, were only approximations. Thus, we assumed that similarity could be represented geometrically, that subjects made continuous discriminations among the concepts (i.e., they did not see them as falling into homogeneous clusters), that they combined differences along different dimensions according to the Euclidean metric, that within-domain and between-domain distance are linear and additive in their effects on comprehensibility and aptness, that subjects rank options in a manner consistent with the Luce choice rule, that their successive rankings can be described by successive applications of this rule, and that the form of the function linking aptness and choice was linear or exponential. Although these assumptions are reasonable, none of them follow directly from a consideration of metaphor; all of them probably introduce some inaccuracy.

Another limitation involves the omission of other factors that clearly affect aptness or comprehensibility, but do not discriminate among the positions of interest. The aptness of metaphor depends on formal criteria (beside comprehensibility), criteria of content (beside similarity), and criteria involving whether the form suits the content. None of these other considerations were taken into account here--whether we can visualize the metaphor, whether its phrasing is rhythmic or metrical, whether the

concepts are pleasant or repulsive, whether the tone and rhythm resonate with the ideas, whether the correspondences involved are fresh or cliched--none were included, but surely all of them affect aptness. And, of course, the single most important determinant of comprehensibility--whether we know the ordinary meanings of the terms used--was also the most obvious determinant but the least relevant to the theories.

The theory and data presented here deal primarily with representational and structural issues in the theory of metaphor, and as such, do not offer an explicit process model of how metaphors are understood and evaluated. A model addressing some of these issues has been offered elsewhere, however (see Sternberg, Tourangeau, & Nigro, in press; Sternberg & Nigro, Note 3), based upon a process theory of reasoning by analogy (Sternberg, 1977a, 1977b). The information-processing model acts upon the representation proposed in this article, and thus may be viewed as complementary to the present work.

The final limitation on these studies involves their generality. Metaphors occur both as literary devices and in everyday language. The metaphors in this study, because they were designed to vary certain factors in an unconfounded and unbiased way, differ from both the literary and ordinary uses of metaphors: these metaphors are clearly worse than literary metaphors and less comprehensible than everyday ones. The problem of generality is always particularly acute for studies with implications for aesthetics. On the one hand, the study may use actual works of art as stimuli and suffer from the problems of confounds among the variables of interest and bias in the selection of works of art; or, on the other

hand, the study may use artificially constructed stimuli (such as ours) that are likely to differ markedly from actual works of art. Here, it seems doubtful that comprehensibility is really unaffected either in literature or in ordinary language by any form of distance between tenor and vehicle. Perhaps our rating-scale measure of comprehensibility was insensitive. Perhaps the form of the relationship between distance and comprehensibility varies with the level of aptness or is some threshold function. Since we studied only metaphors relatively low in aptness and comprehensibility, we are unable to test these possibilities.

Conclusions

Despite these limitations, some conclusions seem warranted. In line with the assumptions of all but the most extreme anomaly theorists, the comprehensibility of a metaphor relates positively to its aptness. Contrary to a view held by most comparison theorists and many anomaly theorists, the key similarity underlying a metaphor is not a pre-existing common category membership; indeed, the evidence of Experiment 1 indicates the more dissimilar the most salient categories of tenor and vehicle are, the better the metaphor. At the extreme, when tenor and vehicle are from a common category, the metaphor is a bad one, even perhaps anomalous. Instead, it is the similarity of the positions of the tenor and vehicle within their separate categories that is central in determining aptness. All of these findings support the domains-interaction view--the view that in metaphor an object within one domain is seen as occupying a position or role similar to that of an object in another domain.

Reference Notes

1. Tourangeau, R., & Sternberg, R. J. Understanding and appreciating metaphors. NR150-412 ONR Technical Report #11. New Haven: Department of Psychology, Yale University, 1978.

2. Kraut, R. Effects of incompatibility and ambiguity on the humor and poetic quality of metaphors. Master's thesis submitted to Yale University.

3. Sternberg, R. J., & Nigro, G. Component processes in metaphoric comprehension and appreciation. Manuscript in preparation, 1978.

References

Aristotle. Poetics. In W. Fyfe (tr.), Aristotle: the poetics. Cambridge, Mass.: Harvard University Press, 1927.

Aristotle. Rhetoric. In L. Cooper (tr.), The rhetoric of Aristotle. New York: Appleton-Century-Crofts, 1932.

Basso, K. "Wise words" of the Western Apache: metaphor and semantic theory. In K. Basso & H. Selby (Eds.), Meaning in anthropology. Albuquerque, New Mexico: University of New Mexico Press, 1976.

Beardsley, M. The metaphorical twist. Philosophy and Phenomenological Research, 1962, 22, 293-307.

Bickerton, D. Prolegomena to a linguistic theory of metaphor. Foundations of Languange, 1969, 5, 36-51.

Billow, R. Metaphor: a review of the psychological literature. Psychological Bulletin, 1977, 84, 81-92.

Black, M. Models and metaphors. Ithaca, N.Y.: Cornell University Press, 1962.

Brooke-Rose, C. A grammar of metaphor. London: Seeker and Warburg, 1958.

Campbell, P. Metaphor and linguistic theory. Quarterly Journal of Speech, 1975, 61, 1-12.

Chomsky, N. Degrees of grammaticalness. In J. Fodor & J. Katz (Eds.), The structure of language. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Gorsuch, R. L. Factor analysis. Philadelphia: Sauders, 1974.

Guenther, F. On the semantics of metaphor. Poetics, 1975, 4, 199-200.

Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw-Hill, 1956.

Henle, P. Metaphor. In P. Henle (Ed.), Language, thought, and culture. Ann Arbor, Michigan: An Arbor Paperbacks, 1966.

Hesse, M. Models and analogies in science. Notre Dame, Indiana: University of Notre Dame Press, 1966.

Katz, J. Semi-sentences. In J. Fodor & J. Katz (Eds.), The structure of language. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Kruskal, J. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27. (a)

Kruskal, J. Nonmetric multidimensional scaling: a numerical method. Psychometrika, 1964, 29, 115-129. (b)

Luce, R. D. Individual choice behavior. New York: Wiley, 1959.

Malgady, R., & Johnson, M. Modifiers in metaphors: effects of constituent phrase similarity on the interpretation of figurative sentences. Journal of Psycholinguistic Research, 1976, 5, 43-52.

Miles, J. Style and proportion. Boston: Little, Brown, and Company, 1967.

Ortony, A. The role of similarity in similes and metaphors. In A. Ortony (Ed.), Metaphor and thought. Cambridge, England: Cambridge University Press, in press.

Ortony, A., Reynolds, R. E., & Arter, J. A. Metaphor: Theoretical and empirical research. Psychological Bulletin, 1978, 85, 919-943.

Percy, W. The message in the bottle. New York: Farrar, Strauss, and Giroux, 1954.

Richards, I. The philosophy of rhetoric. Oxford: Oxford University Press, 1971 (1936).

Rips, L., Shoben, E., & Smith, E. Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 1-21.

Rumelhart, D., & Abrahamson, A. A. A model for analogical reasoning. Cognitive Psychology, 1973, 5, 1-28.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 1962, 27, 125-140. (a)

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika, 1962, 27, 219-246. (b)

Sternberg, R. J. Component processes in analogical reasoning. Psychological Review, 1977, 84, 353-378. (a)

Sternberg, R. J. Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, N. J.: Erlbaum, 1977. (b)

Sternberg, R. J., Tourangeau, R., & Nigro, G. Metaphor, induction and social policy: The convergence of macroscopic and microscopic views. In A. Ortony (Ed.), Metaphor and thought. Cambridge, England: Cambridge University Press, in press.

Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

Van Dijk, T. Formal semantics of metaphorical discourse. Poetics, 1975, 4, 173-198.

Wheelwright, P. Metaphor and reality. Bloomington, Indiana: Indiana University Press, 1962.

Footnotes

This work has benefited from conversations of the authors with John Anderson, Ruth Day, Michael Gardner, Dedre Gentner, John Goldin, William McGuire, Georgia Nigro, and Andrew Ortony. We acknowledge in particular the seminal ideas of Phoebe Ellsworth and Amos Tversky, whose influences have critically shaped the present research. Robert Abelson, Thomas Berndt, Irvin Child, and Michael Kubovy commented extensively on earlier drafts of this article, and the final version reflects their careful attention. This research was supported in part by Contract N0001478C0025 from the Office of Naval Research to Robert J. Sternberg, and was done while Roger Tourangeau was supported by an NIMH graduate traineeship. Requests for reprints should be sent to Roger Tourangeau, Department of Psychology, Connecticut College, New London, Connecticut, 06320, or to Robert J. Sternberg, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, Connecticut 06520.

[1]In making this assignment, we implicitly assumed that the between-domain dimensions were unrelated (orthogonal) to the within-domain dimensions. We further assumed that all members of a domain had identical values on the between-domain dimensions; we assumed, for example, that all the historical figures were equally human. Results from Rips, Shoben, and Smith (1973), however, indicate that at least in some cases, this assumption may not hold. Because the between-domain coordinates (which were factor loadings) were on a different scale from the within-domain coordinates (which were sums of means on nine-point ratings scales), in computing overall distance, we standardized the coordinates to have the same means and variances.

[2]The selection process also involved several other constraints. For each domain, a tenor was randomly drawn from each quadrant. Each of these tenors was paired with one of the four domains closest to the domain itself. Each of these four domains had been randomly paired with a quadrant; a vehicle was drawn randomly from that quadrant of that domain. No tenor was paired with itself as a vehicle.

Table 1

Factor Loadings for Birds

Variable	Factor 1 (Power-Aggression)	Factor 2 (Prestige)
Strong	.96	-.02
Powerful	.98	.01
Vigorous	.89	.20
Effective	.92	.11
Controlling	.94	.13
Forceful	.98	.00
Aggressive	.97	-.15
Fierce	.96	-.08
Warlike	.94	-.15
Predatory	.95	.10
Dangerous	.94	-.11
Hunting	.93	.05
Admired	-.15	.98
Noble	.05	.99
Admirable	-.06	.99
Esteemed	-.07	.98
Honored	.16	.95
Respected	.11	.98

Table 2

Correlations Among Factor Loadings

Power/Aggression Factor

Domain	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) U.S. Figures	---							
(2) World Leaders	.89	---						
(3) Land Animals	.97	.90	---					
(4) Birds	.96	.96	.98	---				
(5) Aircraft	.95	.95	.92	.95	---			
(6) Land Vehicles	.96	.96	.98	.99	.95	---		
(7) Fish	.95	.97	.96	.99	.97	.99	---	
(8) Ships	.92	.92	.96	.97	.87	.97	.94	---

Prestige Factor

Domain	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) U.S. Figures	---							
(2) World Leaders	.57	---						
(3) Land Animals	.95	.60	---					
(4) Birds	.80	.80	.88	---				
(5) Aircraft	.52	.95	.60	.86	---			
(6) Land Vehicles	.88	.80	.92	.98	.83	---		
(7) Fish	.78	.94	.83	.94	.92	.95	---	
(8) Ships	.74	.51	.84	.90	.65	.88	.77	---

Table 3

Factor Loading of Domains

Variable	Factor 1 (Animals)	Factor 2 (Vehicles)	Factor 3 (Human)
Aircraft	-.03	.98	-.02
Land Vehicles	.01	.90	.19
Ships	.38	.82	-.19
Birds	.98	-.11	-.13
Land Mammals	.80	.24	.05
Aquatic Animals	.86	.09	-.16
Modern World Leaders	-.42	-.06	.84
U.S. Historical Figures	.49	.13	.56

Table 4

Correlations Among Scales: Experiment 1A

	Good	Interesting	Apt	Like	Quality
Good-Bad		.86	.98	.94	.98
Interesting-Dull			.82	.89	.92
Apt-Not Apt				.92	.97
Like-Dislike					.97
Quality					

Table 5

Correlations of Distances and Quality

	Within-Domain	Between-Domain	Overall
Quality	-.39**	.27*	-.01
Good	-.40**	.24	-.03
Interesting	-.30*	.35**	.10
Apt	-.37**	.20	-.07
Like	-.39**	.25*	-.01

*p < .05 (two-tailed)

**p < .01 (two-tailed)

Table 6

Distances on Individual Dimensions and Quality

	Quality	Good	Interest	Apt	Like
Within-Domain					
Aggression/Power	-.32**	-.32**	-.28*	-.34**	-.29*
Prestige	-.23	-.24	-.13	-.24	-.25*
Between-Domain					
Animal	.29*	.26*	.35**	.22	.32**
Vehicle	-.03	-.04	-.02	-.04	-.06
Human	.30*	.23	.38**	.26*	.25*

* $p < .05$ (two-tailed)

** $p < .01$ (two-tailed)

Table 7

Correlation of Distance and Quality

Without "Anomalous" Metaphors

	Between-Domain	Within-Domain
Quality	.17	-.44**
Good	.15	-.45**
Interesting	.26*	-.34*
Apt	.14	-.46**
Like	.13	-.42**

* $p < .05$ (two-tailed)

** $p < .01$ (two-tailed)

Table 8

Correlations of Comprehensibility and Quality

	Easy	Fast	Comprehensible
Quality	.66	.52	.63
Good	.67	.52	.64
Interesting	.56	.43	.53
Apt	.68	.51	.64
Like	.62	.52	.61

Note: All correlations were significant, $p < .01$, two tails.

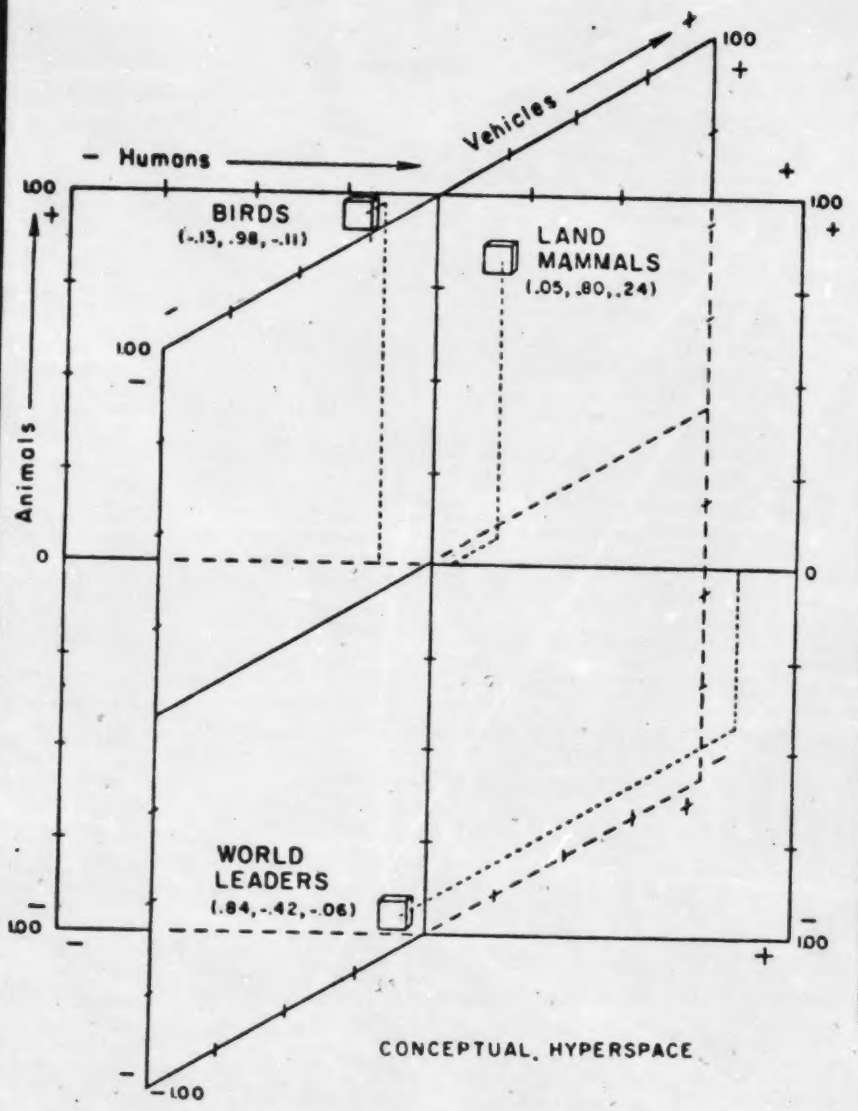
Table 9

Multiple Regression on Quality

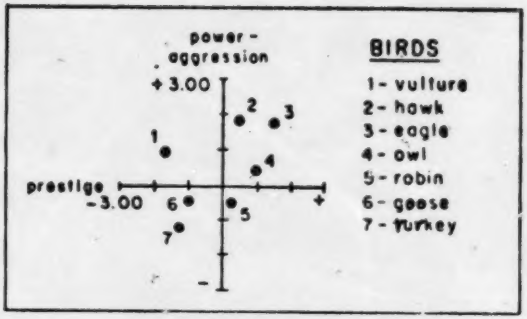
Predictor	Simple r	beta	F
Within distance	-.39	-.40	12.9**
Between distance	.27	.28	6.2*
Comprehensibility	.63	.59	49.0**

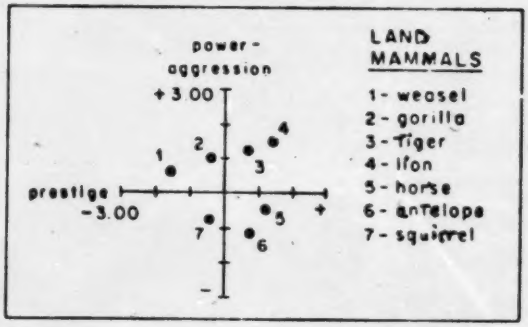
* $p<.05$

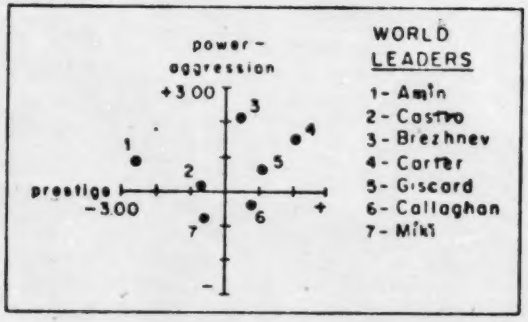
** $p<.01$



CONCEPTUAL HYPERSPACE





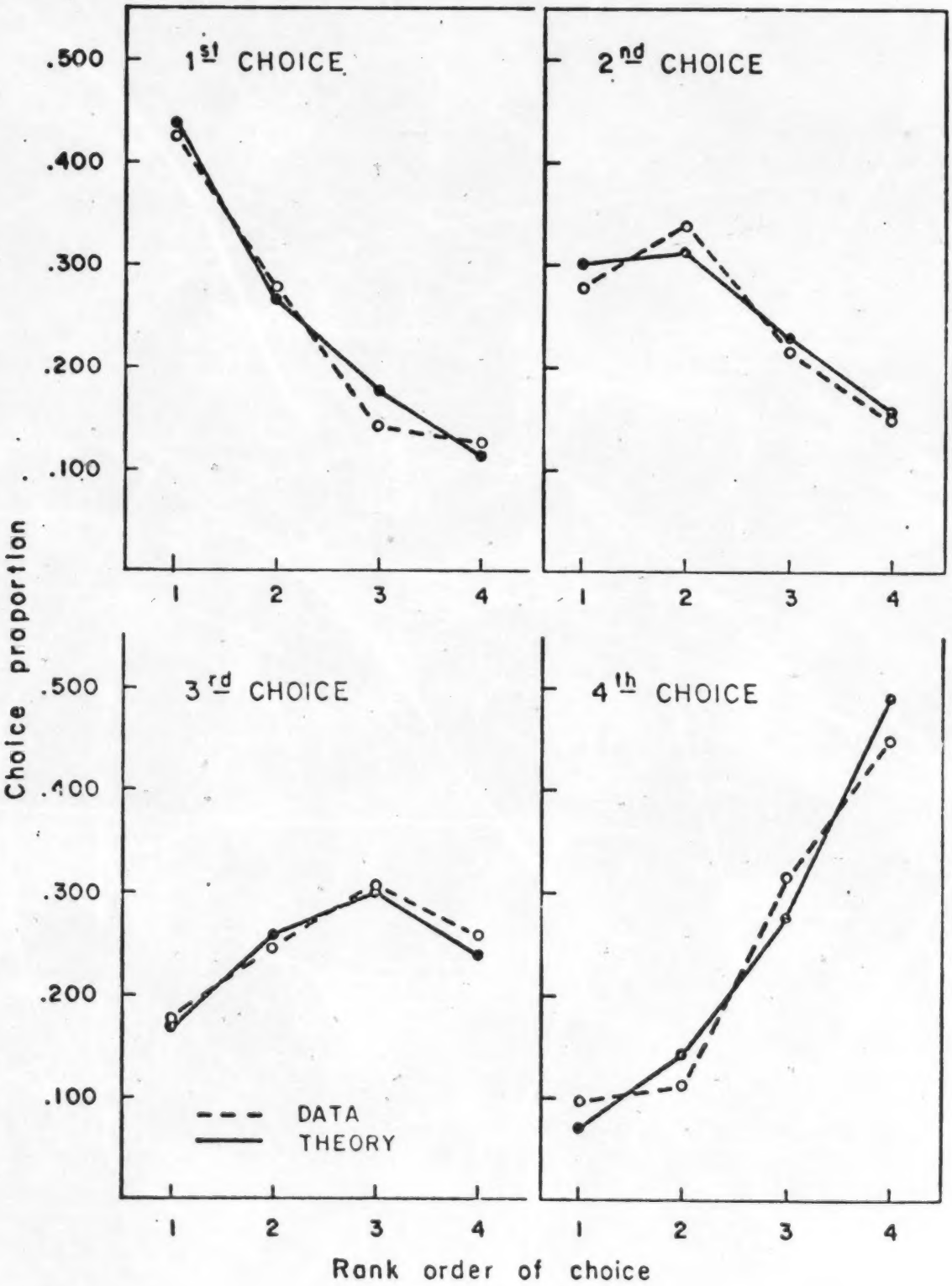


LOWER ORDER SPACE

Figure Captions

Figure 1. Relation of higher-order and lower-order spaces. To the left is shown the space of domains, or hyperspace. Each point in this space is itself a full space of a lower-order, as shown to the right. For example, the point "birds" in the hyperspace maps into the space of birds at the top right of the figure.

Figure 2. Fit of the exponential model with aptness as the predictor.


1st CHOICE
2nd CHOICE
3rd CHOICE
4th CHOICE
.500
.400
.300
.200
.100
1
2
3
4
Choice proportion
DATA
THEORY
Rank order of choice